AccessMED, a Tool for Browsing, Searching and Maintaining a Large Controlled Medical Vocabulary

Barry A. Allen, PhD; Randolph C. Barrows Jr., MD; Nilesh Desai, MS

Department of Medical Informatics

Columbia University, New York, NY

We demonstrate AccessMED, an X Windows graphical user interface that supports development and maintenance of a large semantically-structured controlled medical vocabulary. AccessMED provides browsing, high-speed searching, and editing functions for a unix-based vocabulary server at Columbia-Presbyterian Medical Center (CPMC). The vocabulary server utilizes a shared-memory model (shmMED) of CPMC's Medical Entities Dictionary (MED) that is optimized for speed and compactness. The MED contains nearly 50,000 entities, each of which is defined with a "slot" set chosen from over 150 different attributes. Slots may be literal (string or numeric) or semantic valued (points to another entity in the dictionary), and each may have multiple instances. In addition, the shmMED includes more than 140,000 words mapped to related terms, and over 8,000 synonymous word and word-fragment relationships. At CPMC the MED has become critical in advanced health information applications, such as coded data storage, event-driven decision support, and intelligent clinical results summaries. AccessMED allows multi-user browsing, searching, and (beta-test) editing capabilities of this vocabulary.

AccessMED is implemented in layers. An extensive C library (MEDlib) interface provides access to the shmMED and is available directly on the vocabulary server or through RPC calls from remote peers. An ASCII menu-based browser, a command line query tool, and a web-based browser have been developed utilizing MEDlib. An API for graphical interaction with vocabulary content and structure (MEDViewer) has also been developed. MEDViewer provides a graphical toolset for composing highly refined lexical and semantic searches, compound (boolean) searches, and an interface for graphically browsing the vocabulary. The MEDViewer API is used within an ambulatory care system (DOP - Decision-supported Outpatient Practice system) for coded data capture by enabling a limited subset of capabilities, including graphical browsing, for naive medical users.

AccessMED utilizes and extends MEDViewer services to provide a mulit-user MED browser and editor for developers and vocabulary domain experts. Lexical searching and graphical navigation techniques are generally both necessary by expert users, and are commonly employed in a complementary fashion. Lexical searches with or without semantic filters are generally used for initial entry into the vocabulary to identify desired concepts. Graphical navigation is generally used to review concept relationships and perform searches within a local domain of vocabulary.

A basic lexical search involves entering words and word fragments (tokens) that represent the concept of interest. All tokens are automatically stemmed, stemmed words are then expanded to include synonymous words, the word set is mapped to entities containing an instance of a word set member, and finally the intersection set of entity mappings is returned. In contrast, a code search returns all entities related to a specified entity via *any* semantic relationship. A powerful technique combines both code and lexical searches.

Additional semantic filtering options allow one to require the presence or absence of specific slots in the entity mapping set, or restrict a slot's literal or semantic value. For example, one may require the HAS_GENERIC_DRUG_NAME attribute to be not only defined, but actually instantiated with a value, for each mapped entity. In addition, each instantiated value can be forced to match (the expanded) search tokens or code values. Further options and search complexity will be demonstrated.

AccessMED enables an extensive set of options to control graphical browsing of the MED as well. The default viewing mode presents an acyclic graph around a focal concept, showing parents, siblings, and children, as determined by the "SUBCLASS_OF / SUPERCLASS_OF" slot pair relationship. Navigation is accomplished as the graph refocuses when selecting any visible node. The graph may also be focused by selecting a concept from the output of a lexical search. One viewing option includes all ancestors back to the root node ("Medical Entity") on a single display. The MED currently has 31 semantic-valued slot pair relationships; another option allows the graphical viewing of any combination of these relationships. Thus "Allen Chem-7" is a subclass-of Chem-7 and has no children. However, if viewing of the "HAS_PARTS / PART_OF" relationship is enabled, Allen Chem-7 shows 9 additional "children" (the seven chemistry test entities and two comment entities that are parts of a CHEM-7).